

Künstliche Intelligenz

Künstliche Intelligenz gilt als das "nächste große Ding" im Zusammenhang mit Digitalisierung. Dass hinter den die Künstliche Intelligenz befördernden Lernmethoden Verfahren stecken, die schon im 18. Jahrhundert entdeckt wurden (Satz von Bayes) oder zu Beginn des 20. Jahrhunderts (Hauptkomponentenanalyse) beschrieben wurden, ist nicht sehr bekannt. Aber der Siegeszug dieser Methoden und Verfahren kann jetzt mit der enorm gestiegenen Leistungsfähigkeit der Halbleiterelektronik in Verbindung mit der Entwicklung hoher Programmiersprachen und Software-Designmethoden stattfinden.

Eine sehr pragmatische Definition bezeichnet Künstliche Intelligenz (KI) als diejenigen Fähigkeiten des Computers auf der Basis maschinellen Lernens um Dinge zu tun, die Menschen derzeit besser können⁵³. Andere, weniger heuristische Definitionen sind:

- KI ist eine Wissenschaftsdisziplin der Informatik, die sich mit der Entwicklung von Hard- und Software Systemen befasst, die Probleme lösen können, für deren Lösung gemeinhin Intelligenz erforderlich ist.
- Als KI werden Techniken bezeichnet, die in Hard- und Softwaresystemen für die Leistungssteigerung eingesetzt werden.
- KI ist ein Hard- und Softwaresystem, welches ein intelligentes Problemlösungsverhalten zeigt.
- Als KI wird ein künstliches Wesen bezeichnet, welches Intelligenz besitzt.

Der Begriff „Künstliche Intelligenz“ ist sehr anfällig für Missverständnisse.

Der Begriff unterstellt, dass die Maschine, die künstliche Intelligenz aufweist, über menschengleiche oder sogar den Menschen übertreffende Fähigkeiten aufweist. Betrachtet man die Rechenleistung, so ist jeder Taschenrechner einem Menschen überlegen. Ein programmierbarer Taschenrechner kann schon als eine Vorstufe für den Einstieg in die KI angesehen werden: er kann zum Lösen strategischer Probleme, das sich als Intelligenz interpretieren lässt, eingesetzt werden. Wenn das System dann noch selbstständig und effizient Probleme löst, kann es als intelligent bezeichnet werden. Der Grad der Intelligenz hängt vom Grad der Selbstständigkeit, dem Grad der Komplexität des Problems und dem Grad der Effizienz des Problemlösungsverfahrens ab und kann über Intelligenzfaktoren⁵⁴ gemessen werden:

- Zahlenrechnen (numbers)
- Sprachverständnis (verbal comprehension)

⁵³ Wolfgang Ertel, „Grundkurs Künstliche Intelligenz“, Springer Verlag, 2016

⁵⁴ Thurstone, Louis Leon (1887-1955); University of Chicago

- Raumvorstellung (space)
- Gedächtnis (memory)
- schlussfolgerndes Denken (reasoning)
- Wortflüssigkeit (word fluency)
- Auffassungsgeschwindigkeit (perceptual speed)

Das Missverständnis besteht nun darin, dass der Begriff „künstlich“ eigentlich beinhaltet, dass der Computer eben kein Mensch ist, nicht aus Fleisch und Blut besteht, kein Schmerzempfinden und keine Sterblichkeit aufweist. Das Empfinden von Leid und Freude müsste ihn, den Computer, von einer künstlichen Maschine zu einem Organismus machen. Da er diese Fähigkeiten eben nicht hat, können KI-Systeme nicht fühlen, sie besitzen keine Instinkte, Stimmungen, Wahrnehmungen und Berührungen als Teil des Bewusstseins.

Macht man sich diesen Unterschied klar, so kann man in die spannende Frage eintauchen, inwieweit KI-Algorithmen heute die Intelligenzfaktoren ausfüllen können.

KI-Technologien

Maschinelles Lernen (ML) wird als Schlüsseltechnologie der Künstlichen Intelligenz verstanden. Maschinelles Lernen bezweckt die Generierung von »Wissen« aus »Erfahrung«, in dem Lernalgorithmen aus Beispielen ein komplexes Modell entwickeln

Viele Methoden des ML sind nicht neu. Schon um 1950 wurden erste Arbeiten zu neuronalen Netzen veröffentlicht. (Künstliche) Neuronale Netze verarbeiten Vektoren in mehrdimensionalen Zahlenräumen, die von Schicht zu Schicht durch die Knoten transformiert werden, um auch komplexere, nicht-lineare Funktionen zu approximieren.

Aber es erfolgte alsbald eine Stagnation der Forschung und Entwicklung (FuE), da neuronale Netze aus zwei wesentlichen Gründen nicht die gewünschten Erfolge brachten: 1. sind damals die Computer zu langsam und 2. erschien vielfach die Plausibilisierung von neuronalen Netzen als schwierig. Über ein Training wird in einem künstlichen neuronalen Netz ein Modell erstellt, das die Eingabesignale mit den Ausgabesignalen ins Verhältnis setzt. Dabei nehmen die Verbindungen zwischen den Neuronen unterschiedliche Gewichtungen an, um für jeden Datensatz eine möglichst zutreffende Klasse zu bestimmen. Dies analytisch zu plausibilisieren ist nicht einfach. Seit dem Jahr 2000 gewinnen die neuronalen Netze jedoch mit der massiv steigenden Rechenleistung ibds. im Zusammenhang mit Deep Learning wieder an Bedeutung.

Seitdem (ca. seit 2005) immer größere Datenmengen aus unterschiedlichen Quellen („BigData“) verarbeitet werden können, können mit ML Technologien neue Anwendungen erschlossen werden. So ist die Muster- und Spracherkennung auf der Basis des überwachten Lernens (supervised learning) u.a. auf der Basis des Satzes von Bayes in Form des „Naiven Bayes“-Algorithmus

bedeutsam für die Fortschritte in der Spracherkennung und Bildverarbeitung. Die Spracherkennung in Autos konnte durch den Einsatz von Trainingsmethoden seit etwa 2010 deutlich verbessert werden. Beim überwachten Lernen liegt zu jedem Trainingsbeispiel gleich die richtige Antwort vor. Wenn etwa handgeschriebene Ziffern erkannt werden sollen, braucht man zu den Zeichen die richtige Ziffer, so dass die Richtigkeit der maschinellen Zuordnung sofort an den Lernalgorithmus zurückgespiegelt werden kann. Alle Trainingsdaten sind hierzu mit der richtigen Ziffer als Label versehen. – Diese Methoden eignen sich gut für Prognose- und Klassifizierungsaufgaben.

Verfahren des unüberwachten Lernens, bei denen zu erkennende Kategorien oder Muster nicht apriori bekannt sind, werden benutzt um Strukturen und Unterschiede in den Daten zu erkennen, um etwa Gruppen (Englisch: Cluster) ähnlicher Beispiele zu finden. Ein Beispiel dafür ist der k-Means Algorithmus, bei dem in einem iterativen Prozess Distanzen von Punkten zu zunächst bezüglich Anzahl und Lage beliebig gewählten Clusterzentren gemessen werden und in einem zweiten Schritt die Clusterzentren unter Maßgabe der Distanzminimierung neu berechnet werden und diese beiden Schritte solange wiederholt werden, bis sich die Clusterzentren nicht mehr verschieben⁵⁵. Dieses Verfahren hat große Bedeutung für die mustererkennende Bildverarbeitung. Eine weitere wichtige Methode des unüberwachten Lernens ist die Hauptkomponentenanalyse⁵⁶. Hier wird in einem hochdimensionalen Datenraum nach Redundanzen gesucht. Die nicht redundanten Informationen werden so linear kombiniert, dass sich neue Hauptachsen eines (meist 2-dimensionalen) Koordinatensystems ergeben, in denen die Originaldaten dargestellt und separiert werden können um Unterschiede zu erkennen. Eine typische Anwendung ist die Segmentierung von Kundendaten. Ein weiterer Lernstil ist das bekräftigende Lernen (reinforcement learning) für Maschinen, die mit ihrer Umgebung interagieren. Dabei nutzen sie Feedback, das sie auf ihre Aktionen von der Umwelt erhalten, um die Erfolgsaussichten der einzelnen Aktionen in den verschiedenen Situationen besser einschätzen zu lernen. Mit neuronalen Netzen, die sowohl durch überwachtes Lernen konditioniert als auch durch bekräftigendes Lernen trainiert werden, gewann das Computerprogramm AlphaGo in 2017 gegen den Weltranglistenersten Ke Jie im Go-Spiel.⁵⁷

Es gibt kognitive Aufgaben, mit denen die KI heute noch ihre Probleme hat. Spiele wie Schach sind zwar komplex, haben jedoch einfache und feste Regeln. Die Gesamtkomplexität eines Gerichtsverfahrens passt nicht in einen Vektor oder ein Array. Und selbst wenn, dann sind die Regeln

⁵⁵ J. B. MacQueen: Some Methods for classification and Analysis of Multivariate Observations. In: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, Band 1. University of California Press, 1967, S. 281–297

⁵⁶ Karl Pearson: On lines and planes of closest fit to a system of points in space. In: The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. Series 6, 2, 1901, S. 559–572

⁵⁷ <https://www.nature.com/news/google-reveals-secret-test-of-ai-bot-to-beat-top-go-players-1.21253>

des juristischen Taktierens und die abstrakten Ziele extrem schwer zu greifen. Die Schlüsseltechnik der Mustererkennung durch Brute-Force-Methoden anzuwenden ist bei solchen Problemen eher nutzlos.

Hier stößt KI an ihre Grenzen, was sich auch nicht durch Rechenleistung und intelligentere Architekturen ändern wird.

Deep-Learning-Methoden arbeiten besonders gut, wenn sehr viele Daten vorhanden sind und die Problemstellung ganz klar abgegrenzt ist. Aber es gibt auch noch viele Herausforderungen – unter anderem: ein echtes Verständnis abstrakter Konzepte, Transfer von Wissen auf neuartige Anwendungsprobleme, Transparenz und Sicherheitsgarantien, die Unterscheidung zwischen zufälligen und logisch sinnvollen Zusammenhängen.

Künstliche Intelligenz ist auf riesige Datenmengen angewiesen. Verfügt sie über diese nicht, so stößt sie an ihre Grenzen. KI kann heute nur explizit auf Grundlage von Daten lernen, jedoch nicht wie wir Menschen implizit. Wir Menschen können Verallgemeinern und Probleme in verschiedene Kontexte setzen. Die KI hingegen kann dies nicht.

Für die schnelle und umfassende Analyse von Daten ist die KI auf hohe Rechenleistungen angewiesen. Rechenleistungen werden zwar immer besser, doch für viele KI-Anwendungen sind sie heute noch nicht gut genug.

Für die KI stehen für viele Anwendungen Daten nicht in entsprechender Qualität und Quantität zur Verfügung. Datensilos müssen erst aufgebrochen und Datenquellen erschlossen werden.

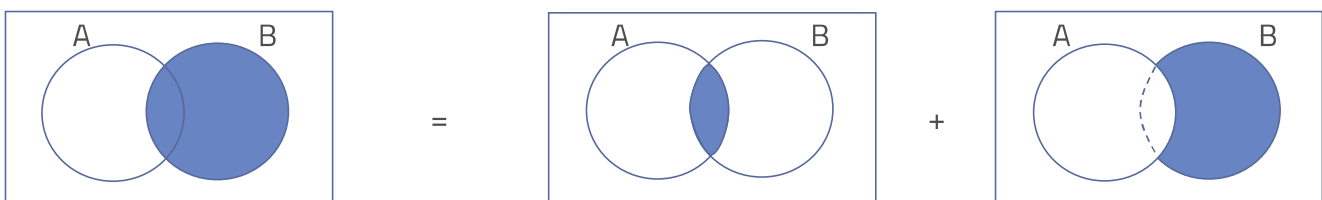


Abb. 17 - Satz von Bayes

Überwachtes Lernen – Satz von Bayes

Der „Naive Bayes“-Algorithmus geht auf den Satz von Bayes zurück⁵⁸. Der Satz von Bayes setzt die bedingte Wahrscheinlichkeit $P(A|B)$, dass ein Ereignis A eintritt unter der Bedingung, dass B eingetreten ist, und die Anfangswahrscheinlichkeit für das Ereignis $P(B)$ in Beziehung mit der umgekehrten bedingten Wahrscheinlichkeit $P(B|A)$ und $P(A)$:

$$P(A|B) * P(B) = P(B|A) * P(A)$$

woraus sofort folgt

⁵⁸ Reverend Thomas Bayes (* 1701/1702, † 1761) war ein englischer Mathematiker, Statistiker, Philosoph und presbyterianischer Pfarrer. Der Artikel „An essay towards solving a problem in the doctrine of chances“ wurde 1763 post-mortem veröffentlicht.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Es besteht also ein Verhältnis zwischen der bedingten Wahrscheinlichkeit $P(A|B)$ und der umgekehrten bedingten Wahrscheinlichkeit $P(B|A)$. Oft sind nur bedingte Wahrscheinlichkeiten (z.B. $P(B|A)$) und die Wahrscheinlichkeit des bedingenden Ereignisses (z.B. $P(A)$) bekannt. Dann kann man sich mit etwas Mengenlehre und dem „Venn-Diagramm“ das Gesetz der totalen Wahrscheinlichkeit klar machen:

Offensichtlich ist die Wahrscheinlichkeit für B gleich der Summe aus der Wahrscheinlichkeit, dass B und A eintritt [geschrieben $p(B \cap A)$] plus der Wahrscheinlichkeit, dass B eintritt und A NICHT [geschrieben $p(B \cap \bar{A})$]! Die Wahrscheinlichkeit, dass B und A eintritt ist nun genau das Produkt aus der bedingten Wahrscheinlichkeit für B unter der Bedingung A multipliziert mit der Wahrscheinlichkeit für A

$$p(B \cap A) = p(B|A) * p(A)$$

Ebenso ist die Wahrscheinlichkeit, dass B eintritt und A nicht [geschrieben $p(B \cap \bar{A})$] gleich dem Produkt aus der bedingten Wahrscheinlichkeit für B unter der Bedingung das A nicht eingetreten ist [geschrieben $p(B|\bar{A})$] multipliziert mit der Wahrscheinlichkeit das A nicht eintritt [geschrieben $p(\bar{A})$]:

$$p(B \cap \bar{A}) = p(B|\bar{A}) * p(\bar{A})$$

Damit folgt für die gesuchte Anfangswahrscheinlichkeit

$$p(B) = p(B|A) * p(A) + p(B|\bar{A}) * p(\bar{A})$$

und der Satz von Bayes kann geschrieben werden als

$$p(A|B) = \frac{p(A) * p(B|A)}{p(B|A) * p(A) + p(B|\bar{A}) * p(\bar{A})}$$

Überwachtes Lernen – Beispiele

Der durch den Satz von Bayes präzierte Zusammenhang zwischen bedingten Wahrscheinlichkeiten ist mathematisch bzw. formal gesehen einfach. Schwieriger hingegen ist das Zusammenfügen des intuitiven Vorverständnisses mit den formalen Ergebnissen und Konzepten. Dies soll an den folgenden drei Beispielen verdeutlicht werden.

Testschärfe

Im Zusammenhang mit medizinischen Tests wird die Sensitivität und Spezifität zur Beschreibung der Testgüte verwendet. Die Sensitivität ist dabei die Wahrscheinlichkeit, dass ein Test Infizierte als infiziert erkennt: der Anteil der Personen mit positivem Testergebnis unter den Infizierten. Die Spezifität ist die Wahrscheinlichkeit, dass er Gesunde als gesund erkennt: der Anteil der Personen mit negativem Testergebnis unter den Nicht Infizierten.

Für die Antigen-Schnelltests auf SARS-CoV-2 wird eine Sensitivität von 80 % und eine Spezifität von 98 % angegeben. Wenn jetzt ein Test positiv ist, wie groß ist dann die Wahrscheinlichkeit, dass die getestete Person tatsächlich Covid-19 hat?

Wir kennen damit zwei bedingte Wahrscheinlichkeiten. Wenn A für das Vorliegen der Infektion steht und B für das positive Testergebnis dann ist $p(B|A) = 0,8$ und $p(\bar{B}|\bar{A}) = 0,98$. Da sich die bedingten Wahrscheinlichkeiten für das Eintreffen eines Ereignisses und das Nicht-Eintreffen zu 1 ergänzen müssen folgt für $p(\bar{B}|A) = 0,2$ und $p(B|\bar{A}) = 0,02$. Vergleichen wir mit der Formel von Bayes, so fehlt noch eine Aussage über die „a-priori-Wahrscheinlichkeit“ von Infektionen $p(A)$. Hier geht entscheidend ein, mit welcher Teststrategie vorgegangen wird: Wird (bei unterstellter moderater Inzidenz von 50) zufällig getestet dann kann unterstellt werden, dass $p(A) \approx \frac{50}{100000} = 0,0005$ und damit $p(\bar{A}) = 0,9995$ ist. Einsetzen in die Formel liefert dann eine Wahrscheinlichkeit von nur $\approx 2\%$, dass bei positivem Testergebnis tatsächlich eine Infektion vorliegt! Allerdings tritt bei 10000 getesteten Personen nur in einem Fall der „worst case“ auf, dass die Person trotz negativem Testergebnis infiziert ist. Wird die Teststrategie geändert, und es wird nicht mehr zufallsorientiert, sondern gezielt auf der Basis von Symptomen getestet und das Vorliegen von Infektionen mit $p(A) \approx \frac{10000}{100000} = 0,1$ unterstellt, dann folgt sofort, dass die Wahrscheinlichkeit für eine Infektion bei positivem Testergebnis auf knapp 82 % steigt! Die Wahrscheinlichkeit, dass das negative Testergebnis falsch ist steigt jetzt auf 2,2 %; das wären bei 10000 getesteten Personen mit negativen Testresultaten immerhin 220 Personen!

Mord

Der Satz von Bayes kann auch im Leben von Juristen eine große Bedeutung haben, was anhand der Daten aus einem Mordprozess dargestellt werden soll.⁵⁹

Es ging dabei um die Frage, ob das Ereignis K (dass der vermeintliche Täter mit dem Opfer Kontakt hatte) stattfand, was der Täter bestritt. An der Hose des Angeklagten wurden Blutspuren der Blutgruppe des Opfers gefunden (Ereignis A). In der zugrunde gelegten Bevölkerung besaßen 15,7 % diese Blutgruppe woraus die bedingte Wahrscheinlichkeit $P(A|\bar{K}) = 0,157$ abgeleitet wurde. Ferner wurden unter den Fingernägeln des Opfers Blutspuren des vermeintlichen Täters gefunden (Ereignis B). Mit der Verteilung für diese Blutgruppe ergibt sich die bedingte Wahrscheinlichkeit $P(B|\bar{K}) = 0,062$. Unter Annahme des unabhängigen Auftretens der Ereignisse A und B ergibt sich für das Zusammentreffen der beiden Ereignisse

$$P(A \cap B|\bar{K}) = P(A|\bar{K}) * P(B|\bar{K}) = 0,009$$

⁵⁹ A.Büchter, H.-W. Henn, „Elementare Stochastik“, Springer-Verlag 2006

Die Wahrscheinlichkeit für den Untersuchungsbefund unter der Voraussetzung, dass kein Kontakt zwischen Angeklagtem und Opfer stattfand, beträgt also 0,9 %.

Der „Gutachter“ errechnete daraus eine Wahrscheinlichkeit von über 99 % für den Kontakt zwischen Opfer und Angeklagtem. Er hatte diese Rechnung jedoch ohne die mathematisch geschulte Richterin gemacht, die wusste, dass verschiedene bedingte Wahrscheinlichkeiten nicht einfach zu eins ergänzt werden dürfen, weil

$$P(A \cap B | \bar{K}) + P(K | A \cap B) \neq 1$$

Die Richterin ordnete weitere Nachforschungen an, die erbrachten, dass der Angeklagte zum Tatzeitpunkt an einem 100 km entfernten Ort weilte!

Klassifizierung von Obstsorten

In einer großen Obstumladestation soll ein Klassifikator zur Erkennung von Bananen, Orangen und anderen nicht näher unterschiedenen Obstsorten angelernt werden. Dazu werden insgesamt 1000 Früchte klassifiziert bezüglich ihrer Länge, ihres Zuckergehalts und ihrer Farbe und das Ergebnis in Form einer Tabelle dargestellt:

Ereignisse	Evidenzen	Lang	Nicht lang	Süss	Nicht Süss	Gelb	Nicht Gelb	Total
Banane		400	100	350	150	450	50	500
Orange		0	300	150	150	300	0	300
andere		100	100	150	50	50	150	200
Total		500	500	650	350	800	200	1000

Abb. 18 - Trainingsdaten Obst-Klassifikator

Aus der Tabelle können die A-priori Wahrscheinlichkeiten ermittelt werden

$$P(\text{Banane}) = 0,5 \quad P(\text{Orange}) = 0,3 \quad P(\text{andere}) = 0,2$$

Es können die Wahrscheinlichkeiten für die Evidenzen ermittelt werden

$$P(\text{Lang}) = 0,5 \quad P(\text{Süss}) = 0,65 \quad P(\text{Gelb}) = 0,8$$

Es können bedingte Wahrscheinlichkeiten angegeben werden

$$P(\text{Lang} | \text{Banane}) = 0,8 \quad P(\text{Lang} | \text{Orange}) = 0 \quad [\text{es gibt keine langen Orangen}]$$

$$P(\text{Gelb} | \text{andere}) = \frac{50}{200} = 0,25 \quad P(\text{Nicht Gelb} | \text{andere}) = \frac{150}{200} = 0,75$$

Jetzt soll der Klassifikator angewandt werden und die erste unbekannte Frucht weist die Merkmale Lang, Süss und Gelb auf. Es ist nun zu prüfen auf Banane, Orange oder andere und die Auswahl der höchsten Wahrscheinlichkeit zur Klassifikation der unbekanntes Frucht auf Basis der 1000-Früchte-Trainingsdaten vorzunehmen.

Es ergibt sich

$$\begin{aligned}
 P(\text{Banane}|\text{Lang, Süss, Gelb}) &= \frac{P(\text{Lang}|\text{Banane}) * P(\text{Süss}|\text{Banane}) * P(\text{Gelb}|\text{Banane}) * P(\text{Banane})}{P(\text{Lang}) * P(\text{Süss}) * P(\text{Gelb})} \\
 &= \frac{0,8 * 0,7 * 0,9 * 0,5}{P(\text{Evidenz})} = \frac{0,252}{P(\text{Evidenz})}
 \end{aligned}$$

$$P(\text{Orange}|\text{Lang, Süss, Gelb}) = 0$$

$$\begin{aligned}
 P(\text{andere}|\text{Lang, Süss, Gelb}) &= \frac{P(\text{Lang}|\text{andere}) * P(\text{Süss}|\text{andere}) * P(\text{Gelb}|\text{andere}) * P(\text{andere})}{P(\text{Lang}) * P(\text{Süss}) * P(\text{Gelb})} \\
 &= \frac{\frac{100}{200} * \frac{150}{200} * \frac{50}{200} * \frac{200}{1000}}{P(\text{Evidenz})} = \frac{0,01875}{P(\text{Evidenz})}
 \end{aligned}$$

und damit mit deutlichem Abstand ($0,252 \gg 0,01875$) eine Klassifikation der Frucht Lang, Süss, Gelb als Banane!

Dieser Algorithmus versucht anhand bedingter Wahrscheinlichkeiten Vorhersagen zu machen und ist auch als „Naiver Bayes“ bekannt, da er die erklärenden Variablen als unabhängig voneinander annimmt. In den Trainingsdaten werden die Wahrscheinlichkeiten geschätzt, mit denen die zu beschreibende Variable einen bestimmten Wert annimmt in Abhängigkeit vom Wert einer beschreibenden Variablen. Eine Vorhersage ergibt sich über das Produkt der Wahrscheinlichkeiten aller beschreibenden Variablen. Trotz des Namens, zeigt sich, dass der Naive Bayes eine gute Performance zeigt, da die Klassifikation keine exakten Wahrscheinlichkeitsschätzungen benötigt, solange die maximale Wahrscheinlichkeit der korrekten Klasse zugewiesen wird!

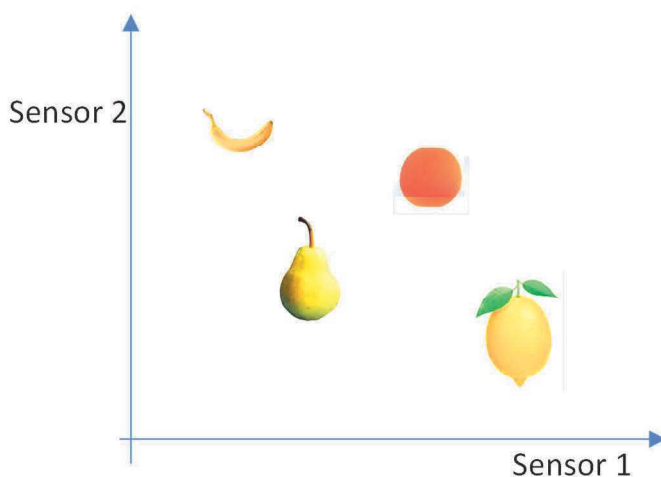


Abb. 19 - Darstellung der Signale zweier bildgebender Sensoren von vier unterschiedlichen Früchten

Unüberwachtes Lernen – Grundidee der Hauptkomponentenanalyse

Häufig wird ein Sachverhalt durch einen mehrdimensionalen Datensatz mit vielen mehr oder weniger stark korrelierten Variablen beschrieben. Mit der Hauptkomponentenanalyse (Principal Component Analysis, PCA) steht ein Verfahren zur Verfügung um mit einer linearen Transformation der n-dimensionalen Variablen in möglichst wenige neue Variablen die relevante Information zu

beschreiben. Das Mass für die Relevanz ist die Varianz. Die neuen Variablen sind orthogonal zueinander und damit unkorreliert. Das Verständnis des mathematischen Verfahrens erfordert gute Kenntnisse der Vektor- und Matrizenrechnung, da eine Hauptachsen-Transformation durchgeführt wird mit einer Minimierung der Korrelation n-dimensionaler Merkmale durch Überführung in einen Vektorraum mit neuer Basis. Ein heuristisches Verständnis der Methode lässt sich aber gut in einem 2-dimensionalen Beispiel gewinnen: Man stelle sich vor es werden jetzt 4 verschiedene Früchte mit zwei verschiedenen bildgebenden Sensoren untersucht. Die Antworten von Sensor 1 und Sensor 2 seien fruchtspezifisch unterschiedlich, ohne Beschränkung der Allgemeinheit wie in der beispielhaften Graphik. - In diesem Beispiel (2 Dimensionen) ist die Situation noch sehr übersichtlich, aber man kann sich leicht vorstellen, dass bei 10 oder gar 100 Sensoren eine graphische Darstellung nicht mehr möglich ist.

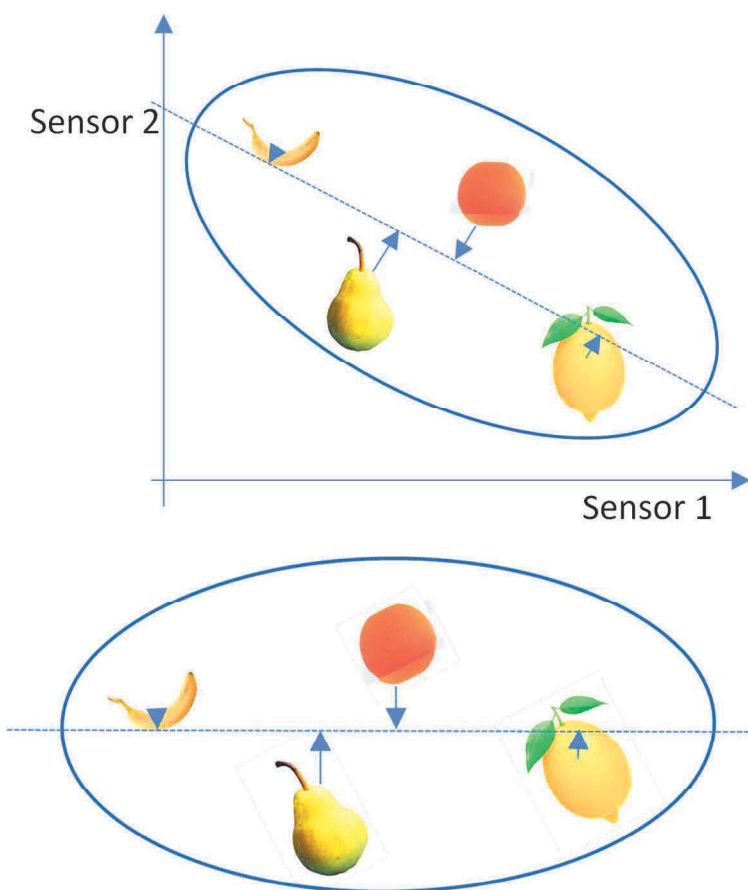


Abb. 20 - Graphische Veranschaulichung der Dimensionsreduktion bei Hauptkomponentenanalysen (oben Konstruktion der Ellipse und erste Hauptachse; unten Koordinaten der ersten Hauptkomponente)

Um die Varianz der Ausdehnung der Früchte im Sensorraum abzuschätzen zeichnet man eine Berandung um die Früchte, die alle Früchte einschließt [Mathematisch wird dazu die Kovarianzmatrix aufgestellt]. Im Allgemeinen wird diese Berandung die Form einer Ellipse haben. Die Ellipse (bzw. das Ellipsoid in einem höherdimensionalen Fall) hat charakteristische Hauptachsen. Die erste Hauptachse beschreibt dabei den größten Anteil der Varianz in den Daten. (Mathematisch

werden die Eigenvektoren und die Eigenwerte der Kovarianzmatrix berechnet um die Lage und Länge der Hauptachsen zu bestimmen). Jetzt müssen nur noch die Koordinaten der Früchte auf die Hauptachse(n) projiziert werden und diese Projektion liefert dann die Lage der Werte auf der ersten Hauptkomponente! Im Beispiel wird deutlich, dass die Koordinatenwerte der Früchte bereits durch Betrachtung der ersten Hauptkomponente unterschieden werden können – hier hat also eine Reduzierung der Dimensionen von einem 2-dimensionalen Raum auf eine Dimension stattgefunden. Mathematisch zusammengefasst heißt das: Die Hauptkomponentenanalyse ist eine statistische Methode der Datenreduktion. Sie basiert auf der Bestimmung der Hauptkomponenten einer Kovarianz- bzw. Korrelationsmatrix. Eine Hauptkomponente ist eine Linearkombination aller beobachteten Variablen. Es gibt so viele Hauptkomponenten wie beobachtete Variablen. Die Hauptkomponenten sind unkorreliert. Die Hauptkomponenten werden so bestimmt, dass sie sukzessive maximale Varianz aufweisen: Die erste Hauptkomponente klärt den größten Anteil der Varianz der beobachteten Variablen auf, die zweite Hauptkomponente erklärt den größten Anteil der nach Extraktion der ersten Hauptkomponente verbleibenden Varianz usw.

	England	Nord-Irland	Schottland	Wales
Alkohol	375	135	458	475
Getränke	57	47	53	73
Fleisch frisch	245	267	242	227
Getreidepro	1472	1494	1462	1582
Käse	105	66	103	103
Süßwaren	54	41	62	64
Fette und Öl	193	209	184	235
Fisch	147	93	122	160
Früchte frisch	1102	674	957	1137
Kartoffel frisch	720	1033	566	874
Gemüse frisch	253	143	171	265
Fleisch ander	685	586	750	803
Gemüse ander	488	355	418	570
Kartoffel ver	198	187	220	203
Gemüse ver	360	334	337	365
Soft drinks	1374	1506	1572	1256
Zucker	156	139	147	175

Abb. 21 - Essverhalten unserer britischen Nachbarn in Gramm/Woche.

Die Stärke der Hauptkomponentenanalyse wird deutlich bei einem Beispiel mit mehr Dimensionen. Betrachtet werde ein Datensatz, der das Essverhalten unserer britischen Nachbarn beschreibt. Die reine Betrachtung des Datensatzes ermöglicht aufgrund der 17 Dimensionen zunächst keine Aussage über die Unterschiede der 4 Länder. Auch die Visualisierung der Zahlenwerte durch die eingezeichneten Balken ändert daran nichts Wesentliches. Verarbeitet man den Datensatz mit Hilfe

der Hauptkomponenten-Analyse⁶⁰ so offenbart der Blick auf die ersten zwei Hauptkomponenten das deutlich abgesetzte Essverhalten der Nordiren, das auch bei einer Betrachtung nur der ersten Hauptkomponente („Reduktion auf eine Dimension“) noch deutlich von den anderen Ländern abweicht.

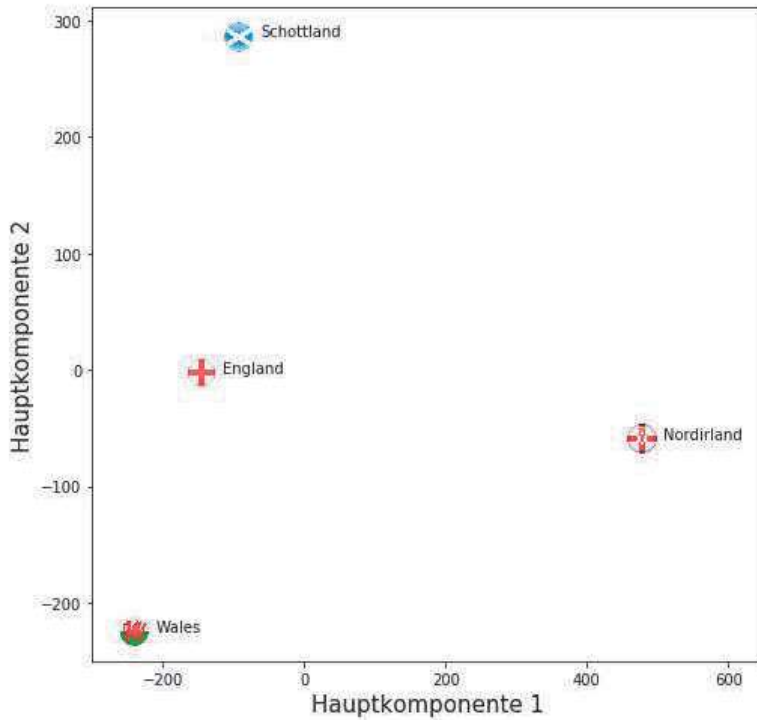


Abb. 22 - Die ersten zwei Hauptkomponenten, die das Essverhalten der Briten charakterisieren.

⁶⁰ Es gibt frei verfügbare Tools, die die HKA implementiert haben; z.B. PAST von der Universität Oslo (<https://www.nhm.uio.no/english/research/infrastructure/past>) oder JASP von der Universität Amsterdam (<https://jasp-stats.org>)